Fiche 002 Niveau : 0 **Gris**

Tirage aléatoire

1 Introduction

L'exercice qui suit s'est présenté naturellement alors que je devais programmer un tirage équirépartie sur un ensemble d'entiers de la forme $\{0, \cdots, n-1\}$. Pour cela je disposais d'un générateur pseudo-aléatoire cryptographique faisant un tirage au sort de manière équiprobable sur un bloc constitué d'un nombre exact d'octets, c'est à dire sur un ensemble de la forme $\{0, \cdots, 2^{8u}-1\}$. Comment à partir d'un tel outil définir un tirage aléatoire équiprobable sur $\{0, \cdots, n-1\}$.
J'ai donc rédigé ce petit exercice qui donne comme solution un algorithme probabiliste (il peut avoir une exécution infini, mais la probabilité d'une telle exécution est nulle et en fait il aboutit pratiquement très vite).

2 Exercice

Soient n et s deux entiers tels que $n < 2^s$. On note P la probabilité équirépartie sur l'ensemble fini $B = \{0, \cdots, 2^s - 1\}$. Soit a un entier appartenant à l'ensemble $A = \{0, \cdots, n-1\}$ et soit :

$$M_a = \{x \in B \mid x \mod n = a\}.$$

1) On suppose dans cette question que l'entier n est lui-même une puissance de 2, c'est-à-dire que $n = 2^k$. Montrer que tous les ensembles M_a sont equiprobables. On peut donc dans ce cas se servir du passage au modulo pour faire un tirage équirépartie sur l'ensemble A des entiers positifs $< n$.

2) On suppose maintenant que n n'est pas une puissance de 2. On peut donc trouver un entier k tel que :

$$2^{k-1} < n < 2^k.$$

a) Donner un exemple de deux ensembles M_a et M_b qui n'ont pas la même probabilité. Donc le passage modulo n ne permet plus de faire un tirage équirépartie sur l'ensemble A des entiers positifs $< n$.

b) Plus généralement montrer que pour toute fonction f de B dans A, les n sous ensembles $f^{-1}(a)$ où a est un élément de A n'ont pas tous le même nombre d'éléments.
Ceci montre que la stratégie suivante pour tirer au sort un élément $a \in A$: on tire au sort un élément $b \in B$ et on calcule $a = f(b)$ ne conduit pas à un tirage équiréparti. En particulier, la solution de proportionnalité qui vient immédiatement à l'esprit et qui consisterait à tirer b au sort et à calculer un arrondi de $\frac{bn}{2^s}$ ne donne pas une probabilité équirépartie sur A.

3) On propose alors pour obtenir à partir d'un tirage équiréparti sur B un tirage équiréparti sur A de procéder de la façon suivante :
On tire un élément de B. S'il est $< n$ on a terminé on a tiré un élément de A au sort. S'il est $\geq n$ on recommence jusqu'à tirer un élément de A.

a) Quelle est la probabilité pour que l'algorithme probabiliste ainsi défini ne se termine pas.

a) A-t-on ainsi un tirage d'un élément de A qui soit de probabilité équirépartie ?

b) Quelle est l'espérance du nombre de tirages à faire ?

c) Peut-on procéder de cette façon si l'exposant s est bien plus grand que l'exposant k.

d) À partir des résultats précédents monter une stratégie pour tirer au sort de manière équirépartie un élément de A, sachant qu'on sait tirer au sort de manière équiprobable un élément de b. Calculer l'espérance du nombre d'itération de la stratégie que vous proposez.

3 Solution

1) Dans cette question, $n = 2^k < 2^s$. Soit $x \in \{0, \cdots, 2^s - 1\}$. En faisant la division euclidienne de x par 2^k on peut écrire x de manière unique sous la forme :

$$x = q\,2^k + r, \quad 0 \leq r < 2^k.$$

Si on fixe r vérifiant $0 \leq r < 2^k$, le nombre d'éléments x tels que $x \mod 2^k = r$ est exactement égal au nombre d'éléments $q \leq$ tels que $0 \leq q\,2^k + r < 2^s$, c'est-à-dire aussi le nombre d'éléments $q \geq 0$ tels que $0 \leq q\,2^k < 2^s$ (en effet si $2^k < 2^s$ alors puisque $r < 2^k$ on a aussi $2^k + r < 2 \times 2^k = 2^{k+1} \leq 2^s$. Ce nombre d'éléments est 2^{s-k} et ne dépend pas de r. En conséquence les 2^k ensembles A_r sont équiprobables. Donc la stratégie qui consiste à tirer de manière équiprobable un élément x dans $\{0, \cdots, 2^s - 1\}$ et prendre comme résultat $x \mod n$ permet, lorsque $n = 2^k$, d'avoir un tirage équiprobable sur $A = \{0, \cdots, n - 1\}$.

2) On suppose maintenant que n n'est pas une puissance de 2. On supposera que $2^{k-1} < n < 2^k$.

a) Considérons les points $a = 0$ et $b = n - 1$ de A. Alors M_a est l'ensemble des multiples de n qui sont dans $B = \{0, \cdots, 2^s - 1\}$. C'est aussi le nombre des entiers $q \geq 0$ tels que $qn < 2^s$. En conséquence le nombre des éléments de M_a est :

$$\#M_a = \left\lfloor \frac{2^s}{n} \right\rfloor.$$

Le nombre des éléments de M_b est quant à lui, le nombre des entiers $q \geq 0$ tels que $qn + n - 1 < 2^s$, c'est-à-dire $(q + 1)n < 2^s + 1$ où encore $(q + 1)n \leq 2^s$. En conséquence :

$$\#M_b + 1 = \left\lfloor \frac{2^s}{n} \right\rfloor.$$

Donc $\#M_b = \#M_a - 1$.

b) Soit f une application de B dans A. Les images réciproques $f^{-1}(\{a\})$ où $a \in A$ forment une partition de B. Donc si ces images réciproques ont toutes le même nombre d'éléments t on a $\#B = t\#A$. Comme $\#A = n$ n'est pas une puissance de 2, ceci n'est pas possible. Donc f ne peut pas être utilisée pour fournir à partir d'un tirage équiprobable d'un élément de B, un tirage équiprobable d'un élément de A.

3) Tirages successifs.

a) La probabilité de tirer un élément $< n$ est $n/2^s$. La probabilité de ne pas tirer un élément $< n$ lors des u premiers tirages est $(1 - n/2^s)^u$. Or $1 - n/2^s < 1$, donc la probabilité de ne jamais tirer un élément $< n$, autrement dit d'avoir une exécution infinie de l'algorithme probabiliste qu'on a mis en place est donc nulle.

b) Si a_1 et a_2 sont deux éléments de A, comme $A \subset B$ et que chaque élément de B a la même probabilité d'être tiré, chaque élément de A a la même probabilité d'être tiré. On a donc un tirage équirépartie sur A.

c) Soit X la variable aléatoire à valeurs dans $\{1, \cdots + \infty\}$ qui à une expérience fait correspondre le nombre de tirage qu'on a dû faire avant de tirer un nombre $< n$. L'espérance mathématique de X est du fait que la probabilité d'une exécution infinie est nulle :

$$E(X) = \sum_{u=1}^{+\infty} p_u u,$$

où p_u est la probabilité pour que l'algorithme s'arrête après le u^{e} tirage, c'est-à-dire :

$$p_u = (1 - \frac{n}{2^s})^{u-1} \frac{n}{2^s}.$$

Donc :

$$E(X) = \sum_{u=1}^{+\infty} u(1 - \frac{n}{2^s})^{u-1} \frac{n}{2^s} = \frac{2^s}{n}.$$

(pour sommer la série qui intervient penser par exemple à la dérivée d'une série géométrique).

d) La formule précédente nous montre que si 2^s est grand devant n, il faudra faire en moyenne beaucoup de tirages

e) Pour éviter cet écueil on peut profiter du résultat de 1) qui dit qu'on peut valablement commencer par prendre un le modulo une puissance de deux. Donc si $2^{k-1} < n < 2^k$ on peut commencer par prendre le modulo 2^k. Ceci ramème le problème au cas où $s = k$. Et dans ce cas

$$E(X) = \frac{2^s}{n} < 2.$$

Voici donc l'algorithme complet dans lequel on suppose $2^{k-1} < n < 2^k$.
On exécute :

1. On tire x au sort dans B.
2. On calcule $y = x \mod 2^k$.

jusqu'à ce que $y < n$.
La moyenne du nombre de tirages nécessaires est plus petite que 2.

Auteur : Robert Rolland
Diffusé par l'Association ACrypTA